UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/082,354	02/26/2002	Brian Jacobsen	38627-170421	2756

26694 7590 05/24/2010
VENABLE LLP
P.O. BOX 34385
WASHINGTON, DC 20043-9998

EXAMINER	
HICKS, MICHAEL J	
ART UNIT	PAPER NUMBER
2165	

MAIL DATE	DELIVERY MODE
05/24/2010	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/082,354	JACOBSEN ET AL.	
	Examiner	Art Unit	
	MICHAEL J. HICKS	2165	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *23 February 2010*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1,8,24,26,28-32,36,38-44,46-48,50,51,54-59 and 62-66* is/are pending in the application.

4a) Of the above claim(s) *54-59* is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1, 8, 24, 26, 28-32, 36, 38-44, 46-48, 50-51, 54-59, and 62-66* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *12/03/2009*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 1, 8, 24, 26, 28-32, 36, 38-44, 46-48, 50-51, 54-59, and 62-66 Pending.

Claims 54-59 Withdrawn.

Claims 2-7, 9-23, 25, 17, 33-35, 37, 45, 49, 52-53, and 60-61 Canceled.

Response to Arguments

2. Applicant's arguments filed 2/23/2010 have been fully considered but they are not persuasive.

As per Applicants arguments regarding the system of Diligenti describing a personal document retrieval system, Examiner respectfully disagree. Examiner notes that while the system of Diligenti may be designed as a personal desktop application, this does not disqualify it as a search engine is not required to be accessed through a web page, but may use a desktop application interface. Further, Examiner notes that while the procedure of Diligenti may include extra steps, such as presenting documents to a user, this additional disclosure does not prevent the Diligenti reference from further disclosing all of the claimed limitations.

As per Applicants arguments regarding the limitation of "filtering, by said search engine, subject specific content of each said object visited to determine a relevance of said subject specific content thereof to said predefined particular subject", Examiner respectfully disagrees. Examiner asserts that the training of the Bayesian classifier

using subject specific content, and the subsequent focused crawling which attempts to crawl only pages relevant to the subject specific training data clearly discloses the claimed limitation. Examiner notes that the abstract of Diligenti make it clear that the classifier is trained with data related to a "specific category". Examiner notes that in order for a page to be determined to be irrelevant, and classified as 'other', it must first be crawled and examined. As such, pages which are determined to be non-relevant are crawled and examined by the focused crawler of Diligenti and subsequently filtered out of the set of paged deemed to be relevant and classified as other. Examiner asserts that the optimization disclosed in Diligenti of pruning the crawl path in order not to pursue paths of documents deemed to be irrelevant does not change the fact that irrelevant pages are indeed crawled and filtered.

As per Applicants arguments regarding the limitation of assigning a weight to each of said words, terms and expressions comprising the subject specific terminology of the lexicon, Examiner respectfully disagrees. Examiner notes that the cited disclosure of Diligenti clearly discloses assigning weights to the words of the lexicon, as noted by Applicant in Applicants remarks. Examiner asserts that the claimed limitation does not indicate that a qualitative weight is assigned, but rather that a weight is assigned to each of the words, terms, and expressions of the lexicon. As such, any weights assigned by the system of Diligenti clearly disclose the limitation.

As per Applicants arguments regarding the cumulative total, Examiner respectfully disagrees. Examiner notes that the disclosure of Section 3.2 make it clear that the TF-IDF vectors which are used to determine the probability for a specific document which is subsequently compared to the threshold is a cumulative total of the 40 highest scoring components from the defined vocabulary (e.g. dictionary). Examiner notes that although subsequent processing is performed on this cumulative total before it is compared to the probability threshold, calculating the cumulative total is an integral step is the process.

As per Applicants arguments regarding the predetermined threshold, Examiner respectfully disagree. Examiner notes that Section 3.3, Paragraphs 3-4 clearly indicates that a confidence threshold is employed to determine relevance of crawled documents. As noted above, Section 3.2 make it clear that the TF-IDF vectors which are used to determine the probability for a specific document which is subsequently compared to the threshold is a cumulative total of the 40 highest scoring components from the defined vocabulary (e.g. dictionary).

As per Applicants arguments regarding "presenting one or more of said components of each of said objects to a human editor via a human computer interface", Examiner respectfully disagrees. Examiner notes that the cited paragraph clearly indicates that a human is relied upon to specify representative websites. Examiner asserts that in order to make a determination that a website is representative, the user

must examiner the website, and as such must be presented with the website. Examiner further asserts that presenting a website to a user includes presenting all aspects of the website, including components of the website to the user.

As per Applicants arguments regarding "permitting the human editor to deem a said object to be a subject specific relevant object", Examiner respectfully disagrees. Examiner notes that the cited paragraph clearly indicates that a human is relied upon to specify representative websites. Examiner asserts that selecting a website that is representative of a subject is clearly equivalent to deeming an object to be a subject specific relevant object. Examiner notes that the intended goal of Diligenti is specifically to allow crawling of focused subject matter. Further, Examiner asserts that regardless of whether Diligenti discloses this method as effective or efficient, it is clearly disclosed.

As per Applicants arguments regarding "permitting the human editor to deem a said object to not be a subject specific relevant object", Examiner respectfully disagrees. Examiner notes that the cited paragraph clearly indicates that a human is relied upon to specify representative websites. Examiner asserts that this selecting, as discussed above, inherently includes the ability of the user to deem a website to be non-relevant, as any site not deemed relevant is inherently deemed non-relevant. Examiner notes that the intended goal of Diligenti is specifically to allow crawling of focused subject matter. Further, Examiner asserts that regardless of whether Diligenti discloses this method as effective or efficient, it is clearly disclosed.

As per Applicants arguments regarding the rejection under USC 103, Examiner respectfully notes that, as Claims 35 and 37 have been canceled, and as the new reference of Menczer et al. ("Adaptive Information Agents in Distributed Textual Environments", Proc. 2nd International Conference on Autonomous Agents, Pages 157-164, 1998 and referred to hereinafter as Menczer) has been included in the rejection to disclose the limitation originally presented in Claims 35 and 37, Applicants arguments regarding these claim are considered to be moot.

In light of the above arguments, and the newly cited reference, the rejection will be updated to reflect amendments made to the claims and maintained.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

4. Claim 62 rejected under 35 U.S.C. 102(b) as being anticipated by Diligenti et al. ("Focused Crawling Using Context Graphs", 26th International Conference on Very Large Databases, Pages 527-534, VLDB 2000 and referred to hereinafter as Diligenti).

As per Claim 62, Diligenti discloses a computer-implemented method of implementing a subject specific search engine to compile and access subject specific information, associated with a predefined particular subject, from a computer network, the method comprising the steps of: traversing links between websites comprising one or more objects on the computer network, by said search engine, the objects respectively comprising at least one of: one or more web pages comprising the websites; and one or components comprising any one or more of said web pages, the objects comprising at least one of: words, terms and expressions (See Page 5, Section 3.3 and Page 7, Column 1, Paragraph 2 which clearly disclose that a search engine may be used to initiate a crawl which traverses links between web pages, wherein the webpages are comprised of words terms and expressions.); filtering, by said search engine, subject specific content of each said object visited to determine relevance of said subject specific content thereof to said predefined particular subject (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject.), wherein said filtering comprises: (a) decomposing said objects into one or more said components (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated.); (b) receiving a lexicon, said lexicon comprising subject specific terminology deemed relevant to the predefined particular subject, the subject specific terminology comprising respective words, terms and expressions (See Page 4, Section 3.2 which clearly discloses that a classifier

vector for a particular subject is computed which includes a vocabulary (e.g. lexicon) associated with that category comprising words terms and expressions.); (c) comparing said decomposed components of said objects to said subject specific terminology of the lexicon to determine whether each said object is a subject specific relevant object (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject.), wherein said comparing comprises: (i) assigning a weight to each of said words, terms and expressions comprising the subject specific terminology of the lexicon (See Page 4, Column 2 and Page 5, Column 1 which clearly disclose that each element of the vocabulary are assigned a weight, at least in that a number of matching words from the vocabulary is determined in the classification calculation (See Equation 3). As such it can be considered that each of the terms are equally weighted.); (ii) if a said word, term or expression comprising the object matches a corresponding said word, term or expression comprising the subject specific terminology of the lexicon, adding a corresponding weight thereof to a cumulative total (See Page 5, Equation 3 which clearly discloses that the weight of each matching element is added to a cumulative total.); and (iii) determining any of said objects to be a subject specific relevant object if the cumulative total surpasses a predefined threshold value (See Page 5, Section 3.3 which clearly discloses that a confidence threshold is employed to determine relevant pages.); (d) based upon said comparing, determining all objects deemed to be subject specific relevant as objects to be passed to a second filter (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.), wherein said second

filter comprises: (aa) presenting one or more of said components of each of said objects to a human editor via a human computer interface (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject. Examiner notes that sites deemed irrelevant (e.g. not meeting the minimum confidence threshold) are categorized as 'other' and not crawled further. Examiner notes Page 2, Column 2, Paragraph 4 which clearly indicates that the process of focused crawling may be directed by a human user, as such Examiner asserts that the above disclosed step can be performed by a human editor via a human computer interface.); (bb) permitting the human editor to deem a said object to be a subject specific relevant object if the human editor determines any of said components comprising said object to be within said predefined particular subject (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject. Examiner notes that sites deemed irrelevant (e.g. not meeting the minimum confidence threshold) are categorized as 'other' and not crawled further. Examiner notes Page 2, Column 2, Paragraph 4 which clearly indicates that the process of focused crawling may be directed by a human user, as such Examiner asserts that the above disclosed step can be performed by a human editor via a human computer interface.); (cc) permitting the human editor to deem a said object to not be a subject specific relevant object if the human editor determines any of said components comprising said object to not be within said predefined particular subject

(See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject. Examiner notes that sites deemed irrelevant (e.g. not meeting the minimum confidence threshold) are categorized as 'other' and not crawled further. Examiner notes Page 2, Column 2, Paragraph 4 which clearly indicates that the process of focused crawling may be directed by a human user, as such Examiner asserts that the above disclosed step can be performed by a human editor via a human computer interface.); and (dd) based upon said (bb) and (cc), determining all objects deemed to be subject specific relevant as objects to be saved (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.); presenting for an indexing operation at said search engine, each object determined to be subject specific relevant to said predefined particular subject based upon said filtering (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.).

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6. Claims 1, 8, 24, 26-34, 36, 38-44, 46-48, 50-51, and 63-66 rejected under 35 U.S.C. 103(a) as being unpatentable over Diligenti et al. ("Focused Crawling Using Context Graphs", 26th International Conference on Very Large Databases, Pages 527-534, VLDB 2000 and referred to hereinafter as Diligenti) in view of Menczer et al. ("Adaptive Information Agents in Distributed Textual Environments", Proc. 2nd International Conference on Autonomous Agents, Pages 157-164, 1998 and referred to hereinafter as Menczer).

As per Claims 1, 40 and 41, Diligenti discloses a computer-implemented method, system, and computer readable medium of implementing a specific search engine to compile and access subject-specific information, associated with a predefined particular subject, from a computer network, the method comprising the steps of: traversing links between websites comprising one or more objects on the computer network, by said search engine, the objects respectively comprising at least one of: one or more web pages comprising the websites; and one or more components comprising any one or more of said web pages, the objects comprising at least one of: words, terms and expressions (See Page 5, Section 3.3 and Page 7, Column 1, Paragraph 2 which clearly disclose that a search engine may be used to initiate a crawl which traverses links between web pages, wherein the webpages are comprised of words terms and expressions.); filtering, by said search engine, subject specific contents of each site said object visited to determine a relevance of said subject specific content thereof to said predefined particular subject (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3

and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject.), and, wherein said filtering comprises: (a) decomposing said objects into one or more said components (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated.); (b) receiving a lexicon, said lexicon comprising subject specific terminology deemed relevant to the predefined particular subject, the subject specific terminology comprising respective words, terms and expressions (See Page 4, Section 3.2 which clearly discloses that a classifier vector for a particular subject is computed which includes a vocabulary (e.g. lexicon) associated with that category comprising words terms and expressions.); (c) comparing said decomposed components of said objects to said subject specific terminology of the lexicon to determine whether each said object is a subject specific relevant object (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject.), wherein said comparing comprises: (i) assigning a weight to each of said words, terms and expressions comprising the subject specific terminology of the lexicon (See Page 4, Column 2 and Page 5, Column 1 which clearly disclose that each element of the vocabulary are assigned a weight, at least in that a number of matching words from the vocabulary is determined in the classification calculation (See Equation 3). As such it can be considered that each of the terms are equally weighted.); (ii) if a said word, term or expression comprising the object matches a corresponding said word, term or expression comprising the subject specific terminology of the lexicon, adding a corresponding weight thereof to a cumulative total

(See Page 5, Equation 3 which clearly discloses that the weight of each matching element is added to a cumulative total.); and (iii) determining any of said objects to be a subject specific relevant object if the cumulative total surpasses a predefined threshold value (See Page 5, Section 3.3 which clearly discloses that a confidence threshold is employed to determine relevant pages.); (d) based upon said comparing, determining all objects deemed to be subject specific relevant as objects to be saved (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.); presenting for an indexing operation at said search engine, each object determined to be site deemed subject specific relevant to said particular subject based upon said filtering (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.); indexing and storing said subject specific relevant objects in a searchable database (See Page 7, Column 2, Paragraph 2 which clearly discloses that the method may be used as a search engine, and as such the results are saved in a searchable database.), and assigning a word score to each word appearing on each subject specific relevant object (See Page 4, Column 2 and Page 5, Column 1 which clearly disclose that each element of the vocabulary is assigned a weight, at least in that a number of matching words from the vocabulary is determined in the classification calculation (See Equation 3). As such it can be considered that each of the terms are equally weighted.);

Diligenti fails to disclose said assigning a word score comprises the steps of: determining all websites found in the database that contain links to the website; for each word on the websites, assigning a word score for that word based at least in part on its presence on each website containing a list to the website; and increasing the word score for each website containing a link to the website when the word appears in close proximity to the link.

Menczer discloses said assigning a word score comprises the steps of: determining all websites found in the database that contain links to the website (See Page 160, Column 1, Paragraph 3, and Column 2, Paragraphs 1-2 which clearly indicate that for each link found in the document is determined and further, that words are weighted based upon the distance they appear from the link in the backlinked page.); for each word on the websites, assigning a word score for that word based at least in part on its presence on each website containing a link to the website (See Page 160, Column 1, Paragraph 3, and Column 2, Paragraphs 1-2 which clearly indicate that for each link found in the document is determined and further, that words are weighted based upon the distance they appear from the link in the backlinked page.); and increasing the word score for each website containing a link to the website when the word appears in close proximity to the link (See Page 160, Column 1, Paragraph 3, and Column 2, Paragraphs 1-2 which clearly indicate that for each link found in the document is determined and further, that words are weighted based upon the distance they appear from the link in the backlinked page.).

It would have been obvious to one skilled in the art at the time of applicants invention to modify the teachings of Diligenti with the teachings of Menczer to include said assigning a word score comprises the steps of: determining all websites found in the database that contain links to the website; for each word on the websites, assigning a word score for that word based at least in part on its presence on each website containing a list to the website; and increasing the word score for each website containing a link to the website when the word appears in close proximity to the link with the motivation to exploit word and link cues to perform distributed tasks on behalf of the user. (Menczer, Abstract).

As per Claim 8, Diligenti discloses discarding all objects determined not to be subject specific relevant based upon said comparing (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed. Examiner notes that the objects deemed not relevant are discarded and not indexed.).

As per Claim 24, Diligenti discloses said filtering the occurs prior to said presenting (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject. Examiner notes that each of these steps occur before prior to results being returned or indexed as the express intent of the above steps is to determine what should be returned or indexed.).

As per Claim 26, Diligenti discloses replacing the lexicon with a lexicon corresponding to a different subject in order to present for said indexing operation create a different set of subject specific relevant objects subject-specific database (See Page 4, Section 3.2 which clearly discloses that multiple classifiers and therefore multiple vocabularies exists.).

As per Claim 28, Diligenti discloses permitting a user to enter a query comprising user-preferred words, terms or expressions, wherein the steps of claim 1 are performed in response thereto (See Page 7, Column 2, Paragraph 2 which clearly discloses that the method may be used as a search engine.).

As per Claim 29, Diligenti discloses displaying information found in said step of searching (See Page 7, Column 2, Paragraph 2 which clearly discloses that the method may be used as a search engine.).

As per Claim 30, Diligenti discloses determining a site ranking for each website associated with information found in said searching step (See Page 7 Column 1, Paragraph 4-Column 2, Paragraph 1 which clearly discloses that the results are ranked and returned to the user.).

As per Claim 31, Diligenti discloses displaying the results of the user query using the site ranking of the information found in the searching step to determine an order in which the results are displayed (See Page 7 Column 1, Paragraph 4-Column 2, Paragraph 1 which clearly discloses that the results are ranked and returned to the user.).

As per Claim 32, Diligenti discloses displaying the results of the user query in a hierarchical format according to the site ranking (See Page 7 Column 1, Paragraph 4-Column 2, Paragraph 1 which clearly discloses that the results are ranked and returned to the user.).

As per Claim 36, Diligenti fails to disclose assigning a word score to each word on the website site based at least in part on how many websites sites linking to the website site also contain the particular word.

Menczer discloses assigning a word score to each word on the website site based at least in part on how many websites sites linking to the website site also contain the particular word (See Page 160, Column 1, Paragraph 3, and Column 2, Paragraphs 1-2 which clearly indicate that for each link found in the document is determined and further, that words are weighted based upon the distance they appear from the link in the backlinked page.).

It would have been obvious to one skilled in the art at the time of applicants invention to modify the teachings of Diligenti with the teachings of Menczer to include assigning a word score to each word on the website site based at least in part on how many websites sites linking to the website site also contain the particular word with the motivation to exploit word and link cues to perform distributed tasks on behalf of the user. (Menczer, Abstract).

As per Claim 38, Diligenti discloses entering a user query; using the user query to search the database (See Page 7 Column 1, Paragraph 4-Column 2, Paragraph 1 which clearly discloses that the results are ranked and returned to the user.); and computing a site ranking for each website site associated with information found in said searching step, the site ranking being computed based on said word scores (See Page 7 Column 1, Paragraph 4-Column 2, Paragraph 1 which clearly discloses that the results are ranked and returned to the user.).

As per Claim 39, Diligenti discloses for each website site associated with information found in said searching step, summing the word scores for that website corresponding to words in the user query (See Page 5, Equation 3 which clearly discloses that

the weight of each matching element is added to a cumulative total. Examiner notes that the user query may be used as the vocabulary.).

As per Claim 42, Diligenti discloses monitoring a depth for each said link, the depth being a reflection of relevance to said predefined particular subject (See Page 5, Column 1, Paragraph 1-2 which clearly discloses that the depth of the link is tracked and taken into account in the relevance judgment.).

As per Claim 43, Diligenti discloses for a given said object site being visited resulting from said link, setting a said depth of any links leading from said object that site to other objects to a depth of a link traversed to reach the given object (See Page 5, Column 1, Paragraph 1-2 which clearly discloses that the depth of the link is tracked and taken into account in the relevance judgment.).; wherein said given object site is determined to be relevant to said predefined particular subject setting the depths of the links leading from said site to zero (See Page 5, Column 1, Paragraph 1-2 which clearly discloses that the depth of the link is tracked and taken into account in the relevance judgment.).; and wherein said given object is determined not to be relevant to said predefined particular subject incrementing the depths of the links leading from said object (See Page 5, Column 1, Paragraph 1-2 which clearly discloses that the depth of the link is tracked and taken into account in the relevance judgment.)..

As per Claim 44, Diligenti discloses comparing the incremented depths to a predetermined maximum depth value (See Page 5, Column 1, Paragraph 1-2 which clearly discloses that the depth of the link is tracked and taken into account in the relevance judgment and that a

maximum depth exists.).; wherein when the incremented depths exceed the predetermined maximum depth value, discarding the links leading from said given object (See Page 5, Column 1, Paragraph 1-2 which clearly discloses that the depth of the link is tracked and taken into account in the relevance judgment and that a maximum depth exists.); wherein when the incremented depths do not exceed the predetermined maximum depth value, traversing one of the links leading from said given objects (See Page 5, Column 1, Paragraph 1-2 which clearly discloses that the depth of the link is tracked and taken into account in the relevance judgment and that a maximum depth exists.).

As per Claim 46, Diligenti discloses a subject specific search engine system operable to compile and permit accessing of subject-specific information, associated with a predefined particular subject, from a computer network, the subject specific search engine system comprising: a memory, connected to a host computer, for storing subject specific relevant objects (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.); the host computer executing software stored upon a computer-readable storage medium, the software comprising: a subject specific smart crawler of said search engine traversing links between websites comprising one or more objects on the computer network, the objects respectively comprising at least one of: one or more web pages comprising the websites; and one or components comprising any one or more of said web pages, the objects comprising at least one of: words, terms and expressions (See Page 5, Section 3.3 and Page 7, Column 1, Paragraph 2 which clearly disclose that a search engine may be used to initiate a crawl which traverses links between web pages, wherein the webpages are comprised of words terms and expressions.); said

subject specific smart crawler performing filtering, a first filter of said search engine, to filter out sites, based on site contents, whose contents are irrelevant to said particular subject, and to permit only sites relevant to said particular subject to pass of subject specific content of each said object visited to determine a relevance of said subject specific content thereof to said predefined particular subject (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject. Examiner notes that sites deemed irrelevant (e.g. not meeting the minimum confidence threshold) are categorized as 'other' and not crawled further.), wherein said filtering comprises: (a) decomposing said objects into one or more said components (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated.); (b) receiving a lexicon, said lexicon comprising subject specific terminology deemed relevant to the predefined particular subject, the subject specific terminology comprising respective words, terms and expressions (See Page 4, Section 3.2 which clearly discloses that a classifier vector for a particular subject is computed which includes a vocabulary (e.g. lexicon) associated with that category comprising words terms and expressions.); (c) comparing said decomposed components of said objects to said subject specific terminology of the lexicon to determine whether each said object is a subject specific relevant object (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further

make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject.), wherein said comparing comprises: (i) assigning a weight to each of said words, terms and expressions comprising the subject specific terminology of the lexicon (See Page 4, Column 2 and Page 5, Column 1 which clearly disclose that each element of the vocabulary are assigned a weight, at least in that a number of matching words from the vocabulary is determined in the classification calculation (See Equation 3). As such it can be considered that each of the terms are equally weighted.); (ii) if a said word, term or expression comprising the object matches a corresponding said word, term or expression comprising the subject specific terminology of the lexicon, adding a corresponding weight thereof to a cumulative total (See Page 5, Equation 3 which clearly discloses that the weight of each matching element is added to accumulative total.); and (iii) determining any of said objects to be a subject specific relevant object if the cumulative total surpasses a predefined threshold value (See Page 5, Section 3.3 which clearly discloses that a confidence threshold is employed to determine relevant pages.); (d) based upon said comparing, determining all objects deemed to be subject specific relevant as objects to be saved (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.); an indexer of said search engine indexing to index the relevant sites the plurality of said objects determined to be subject specific relevant to said particular subject based upon said filtering (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.); and an assigner of said search engine for assigning a word score to each word appearing on each subject specific relevant object (See Page 4, Column 2 and Page 5, Column 1 which clearly disclose that each element of the vocabulary is assigned a weight, at least in that a number of matching words from the vocabulary is determined in the

classification calculation (See Equation 3). As such it can be considered that each of the terms are equally weighted.).

Diligenti fails to disclose said assigning a word score comprises the steps of: determining all websites found in the database that contain links to the website; for each word on the websites, assigning a word score for that word based at least in part on its presence on each website containing a list to the website; and increasing the word score for each website containing a link to the website when the word appears in close proximity to the link.

Menczer discloses said assigning a word score comprises the steps of: determining all websites found in the database that contain links to the website (See Page 160, Column 1, Paragraph 3, and Column 2, Paragraphs 1-2 which clearly indicate that for each link found in the document is determined and further, that words are weighted based upon the distance they appear from the link in the backlinked page.); for each word on the websites, assigning a word score for that word based at least in part on its presence on each website containing a link to the website (See Page 160, Column 1, Paragraph 3, and Column 2, Paragraphs 1-2 which clearly indicate that for each link found in the document is determined and further, that words are weighted based upon the distance they appear from the link in the backlinked page.); and increasing the word score for each website containing a link to the website when the word appears in close proximity to the link (See Page 160, Column 1, Paragraph 3, and Column 2, Paragraphs 1-2 which clearly indicate that for each link found in the document is determined and further, that words are weighted based upon the distance they appear from the link in the backlinked page.).

It would have been obvious to one skilled in the art at the time of applicants invention to modify the teachings of Diligenti with the teachings of Menczer to include

said assigning a word score comprises the steps of: determining all websites found in the database that contain links to the website; for each word on the websites, assigning a word score for that word based at least in part on its presence on each website containing a list to the website; and increasing the word score for each website containing a link to the website when the word appears in close proximity to the link with the motivation to exploit word and link cues to perform distributed tasks on behalf of the user. (Menczer, Abstract).

As per Claim 47, Diligenti discloses said filtering is performed by a first lexicon based filter (See Page 4, Section 3.2 which clearly discloses that a classifier vector for a particular subject is computed which includes a vocabulary (e.g. lexicon) associated with that category comprising words terms and expressions.).

As per Claim 48, Diligenti discloses the lexicon is stored on an interchangeable computer-readable storage medium (See Page 4, Section 3.2 which clearly discloses that a classifier vector for a particular subject is computed which includes a vocabulary (e.g. lexicon) associated with that category comprising words terms and expressions. Examiner notes that the classifiers, including the associated vocabularies are saved.).

As per Claim 50, Diligenti discloses the system further comprises a human-computer interface, and comprises: device for presenting said subject specific relevant objects received from the smart crawler to a human editor via the human-computer interface (See Page 7, Column 2, Paragraph 2 which clearly discloses that the method may be used as

a search engine, and as such the results are saved in a searchable database. Further, See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed. Examiner notes Page 2, Column 2, Paragraph 4 which clearly indicates that the process of focused crawling may be directed by a human user, as such Examiner asserts that the above disclosed step can be performed by a human editor via a human computer interface.); and device for receiving input from the human editor, entered via the human-computer interface, regarding whether to index and store said subject specific relevant objects in the memory (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed. Examiner notes Page 2, Column 2, Paragraph 4 which clearly indicates that the process of focused crawling may be directed by a human user, as such Examiner asserts that the above disclosed step can be performed by a human editor via a human computer interface.).

As per Claim 51, Diligenti discloses at least a second filter performing one or more operations of the first filter (See Page 5, Section 3.3 which clearly discloses that several classifiers may be utilized during the crawling.).

As per Claim 63, Diligenti discloses a computer-implemented method of implementing a subject specific search engine to compile and access subject specific information, associated with a predefined particular subject, from a computer network, the method comprising the steps of: traversing links between websites comprising one or more objects on the computer network, by said search engine, the objects respectively comprising at least one of: one or more web pages comprising the websites; and one or components comprising any one or more of said web pages, the

objects comprising at least one of: words, terms and expressions (See Page 5, Section 3.3 and Page 7, Column 1, Paragraph 2 which clearly disclose that a search engine may be used to initiate a crawl which traverses links between web pages, wherein the webpages are comprised of words terms and expressions.); filtering, by said search engine, subject specific content of each said object visited to determine relevance of said subject specific content thereof to said predefined particular subject (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject.), wherein said filtering comprises (a) decomposing said objects into one or more said components (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated.); (b) receiving a lexicon, said lexicon comprising subject specific terminology deemed relevant to the predefined particular subject, the subject specific terminology comprising respective words, terms and expressions (See Page 4, Section 3.2 which clearly discloses that a classifier vector for a particular subject is computed which includes a vocabulary (e.g. lexicon) associated with that category comprising words terms and expressions.); (c) comparing said decomposed components of said objects to said subject specific terminology of the lexicon to determine whether each said object is a subject specific relevant object, wherein a said object is deemed to be a subject specific relevant object if at least one component thereof matches a corresponding subject specific terminology of the lexicon (See Page 5, Section 3.3 which clearly discloses that for each retrieved and linked page, a reduced vector representation is calculated, Page 4, Column 2, Paragraph 3 clearly indicates that the reduced vector

representation is comprised of components (e.g. words terms and expressions) from the object, and Page 5, Section 3.3 and Page 4, Column 2, Paragraph 3 further make it clear that this vector is compared against a classifier vector to determine relevancy for the object to a predefined particular subject.); (d) based upon said comparing, determining all objects deemed to be subject specific relevant as objects to be saved (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.); presenting for an indexing operation at said search engine, each object determined to be subject specific relevant to said predefined particular subject based upon said filtering (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.); indexing and storing said subject specific relevant objects in a searchable database (See Page 7, Column 1, Paragraph 2 which clearly discloses that the objects deemed relevant are saved and indexed.); and assigning a word score to each word appearing on each subject specific relevant object (See Page 4, Column 2 and Page 5, Column 1 which clearly disclose that each element of the vocabulary is assigned a weight, at least in that a number of matching words from the vocabulary is determined in the classification calculation (See Equation 3). As such it can be considered that each of the terms are equally weighted.).

Diligenti fails to disclose said assigning a word score comprises the steps of: determining all websites found in the database that contain links to the website; for each word on the websites, assigning a word score for that word based at least in part on its presence on each website containing a list to the website; and increasing the word score for each website containing a link to the website when the word appears in close proximity to the link.

Menczer discloses said assigning a word score comprises the steps of: determining all websites found in the database that contain links to the website (See Page 160, Column 1, Paragraph 3, and Column 2, Paragraphs 1-2 which clearly indicate that for each link found in the document is determined and further, that words are weighted based upon the distance they appear from the link in the backlinked page.); for each word on the websites, assigning a word score for that word based at least in part on its presence on each website containing a link to the website (See Page 160, Column 1, Paragraph 3, and Column 2, Paragraphs 1-2 which clearly indicate that for each link found in the document is determined and further, that words are weighted based upon the distance they appear from the link in the backlinked page.); and increasing the word score for each website containing a link to the website when the word appears in close proximity to the link (See Page 160, Column 1, Paragraph 3, and Column 2, Paragraphs 1-2 which clearly indicate that for each link found in the document is determined and further, that words are weighted based upon the distance they appear from the link in the backlinked page.).

It would have been obvious to one skilled in the art at the time of applicants invention to modify the teachings of Diligenti with the teachings of Menczer to include said assigning a word score comprises the steps of: determining all websites found in the database that contain links to the website; for each word on the websites, assigning a word score for that word based at least in part on its presence on each website containing a list to the website; and increasing the word score for each website containing a link to the website when the word appears in close proximity to the link with the motivation to exploit word and link cues to perform distributed tasks on behalf of the user. (Menczer, Abstract).

As per Claim 64, Diligenti discloses indexing the totality of objects determined to be subject specific relevant to yield a subcategory of objects (See Page 3, Section 3 and Page 6, Column 2, Paragraphs 4 which clearly disclose that multiple classifiers exists which are defined for multiple categories.).

As per Claim 65, Diligenti discloses the objects are websites, the computer network comprises the Internet, and the subcategory of objects comprises a portion of the Internet (Internet') (See Page 7, Column 2, Paragraph 2 which clearly discloses that the method may be used as a search engine, and as such the results are saved in a searchable database.).

As per Claim 66, Diligenti discloses performing a searching operation upon the Internet' (See Page 7, Column 2, Paragraph 2 which clearly discloses that the method may be used as a search engine, and as such the results are saved in a searchable database.).

Conclusion

7. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within

TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Points of Contact

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael J. Hicks whose telephone number is (571) 272-2670. The examiner can normally be reached on Monday - Friday 9:00a - 5:30p.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Neveen Abel-Jalil can be reached at (571)272-4074. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Michael J Hicks/
Examiner, Art Unit 2165
Phone: (571) 272-2670
Fax: (571) 273-2670

/Neveen Abel-Jalil/
Supervisory Patent Examiner, Art Unit 2165